DOCUMENT RESUME

ED 310 140 TM 013 723

AUTHOR	van der Linden, Wim J.
TITLE	The Use of Moment Estimators for Mixtures of Two Binomials with One Known Success Parameter. Twente Educational Report Number 10.
INSTITUTION	Twente Univ., Enschede (Netherlands). Dept. of Education.
PUB DATE	Nov 80
NOTE	24p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Equations (Mathematics); *Error of Measurement; *Estimation (Mathematics); Foreign Countries; *Guessing (Tests); *Mastery Tests; Mathematical Models; Monte Carlo Methods
IDENTIFIERS	*Moment Estimators; Parametric Analysis

ABSTRACT

Occasionally, situations arise where mixtures of two binomials with one known success parameter are met. An example in educational testing is the mastery or random guessing model in which an examinee is supposed either to master the items or not to master them and to guess blindly. This paper gives moment estimators for such mixtures and presents results from a Monte Carlo investigation into their statistical properties. The results suggest excellent estimators that can safely be used in most instances. It also indicates how the properties of these estimators relate to those of moment estimators for the case in which both success parameters are unknown. Finally, it is pointed out that in situations in which errors in specifying the true value of the known parameter may occur, it might be prudent to consider this parameter as unknown and to estimate accordingly. (Four data tables are included.) (Author)

TWENTE EDUCATIONAL REPORT
NUMBER 10

THE USE OF MOMENT ESTIMATORS FOR MIXTURES OF TWO BINOMIALS WITH ONE KNOWN SUCCESS PARAMETER

W.J. van der Linden

TWENTE UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF EDUCATIONAL TECHNOLOGY
P.O. BOX 217
7500 AE ENSCHEDE THE NETHERLANDS

COLOFON :

Publisher : Department of Educational Technology
Twente University of Technology

Typist : Paula Achterberg

Reproduction : Central Reproduction Department of the University

Edition : 60

The Use of Moment Estimators for Mixtures of Two Binomials with One Known Success Parameter.

W.J.van der Linden

Twente Educational Report No. 10. Enschede,
The Netherlands: Twente University of Technology,
Department of Educational Technology, November 1980

Abstract

Occasionally, situations arise where mixtures of two binomials with one known success parameter are met. An example in educational testing is the mastery or random guessing model in which an examinee is supposed either to master the items or not to master them and to guess blindly. This paper gives moment estimators for such mixtures and presents results from a Monte Carlo investigation into their statistical properties. The results suggest excellent estimators which can safely be used in most instances. It also indicates how the properties of these estimators relate to those of moment estimators for the case both success parameters are unknown. Finally, it is pointed out that in situations in which errors in specifying the true value of the known parameter may occur, it might be prudent to consider this parameter as unknown and to estimate accordingly.

The Use of Moment Estimators for Mixtures of
Two Binomials with One Known Success Parameter

In some areas of scientific research frequency distributions can be encountered which may be conceived of as mixtures of more elementary distributions. These mixtures usually arise because the frequences were generated by separate random processes each accounting for a different part of the data, or because the data were aggregated across some other, mostly unobserved variable. Typically, mixtures display a multimodal form of which each mode can be identified with one of the underlying distributions. Examples of mixtures can be found in the areas of testing the protective power of sera (Muench, 1936, 1938), classification in anthropology (Rao, 1948), and genetics (Powers, 1951). A review of the use of mixtures is given in Blischke (1963).

An important class of mixtures of distributions in educational and psychological testing is the beta-binomial model, introduced in this area by Keats and Lord (1962). In this model, which is used most when item sampling is involved, the conditional distributions of observed scores given the relative (generic) true score is the binomial; the marginal distribution of the observed scores is the beta-binomial or negative hypergeometric, obtained from the binomials by using a beta distribution over the true score as mixing distribution. Another example can be found in the area of mastery testing (Meskauskas, 1976; van der Linden, 1980a), where a class of models is in use known as latent class or state models (Besel, 1973; Dayton & Macready, 1976; Emrick, 1971; Emrick & Adams, 1969; Macready & Dayton, 1977). For a domain of test items, these models postulate two latent states, a "mastery" and a "nonmastery" state, which are each characterized by a different set of success probabilities for the items. In a simple version of these models, these probabilities are assumed to be equal to each other for the mastery as well as the nonmastery state, implying a mixture of two binomials for the observed score distribution with the probability of a latent master as mixing parameter. Another version allows these probabilities to vary across the items, however, and is basically a mixture of two compound binomial distributions (which, to complicate matters, themselves are mixtures of Bernoulli distributions; see, e.g., Walsh, 1953, 1959, 1963).

In this paper, we focus on mixtures of two binomials with one known success parameter. We shall show how moment estimators can be obtained for the remaining unknown parameters of such mixtures and present results from a Monte Carlo study carried out to explore the statistical properties of these estimators. The latter also involves the properties of an estimated Bayes rule that can be computed from these estimators, and which is suited for classifying observations into the two unknown states represented by the binomial success parameters. As an example of the use of a mixture of two binomials with one known success parameter, we shall refer to the latent class model of mastery testing proposed by Emrick and Adams (1969) in which the success parameter for the nonmasters is set equal to the reciprocal of the number of alternatives of the items. This "mastery or random guessing model" was considered by Reulecke (1977a) and alluded to in an example by Emrick (1971). But readers not familiar with mastery testing could also think of one of the other areas referred to above, for example, mixtures of distributions in genetics in which one of the underlying distributions follows from genetic theory.

Before turning to the derivation of moment estimators for the present case, we note that moment estimators for the case of a known mixing parameter have been considered by Blischke (1962) and that Monte Carlo results for the case of both success parameters as well as the mixing parameter unknown were presented in an earlier paper (van der Linden, 1980b).

Moment Estimators with One Known Success Parameter

Let $X_1, \ldots, X_m$ be m identically and independently distributed random variables, each with possible values $x = 0, 1, \ldots, n$ and probability function

$$(1) \qquad p(x) = \binom{n}{x}\left[(1 - \mu)\alpha^x(1 - \alpha)^{n-x} + \mu\beta^x(1 - \beta)^{n-x}\right].$$

We assume that α is the known parameter and, without loss of generality, that

$$(2) \qquad \alpha < \beta.$$

Note that the strict inequality in (2) avoids degeneration of (1) into a single binomial. For this reason it is also assumed that $\alpha > 0$, $\beta < 1$, and

$$(3) \qquad 0 < \mu < 1.$$

It can be shown that a Bayes rule for deciding whether an observed value x was generated by a process with α or β as value for the success parameter takes a monotone shape. According to this rule we decide on parameter value α for observed values smaller than

$$(4) \qquad c^* = \frac{n \ln \dfrac{1-\beta}{1-\alpha} + \ln \dfrac{\lambda\mu}{1-\mu}}{\ln \dfrac{\alpha(1-\beta)}{\beta(1-\alpha)}},$$

and β otherwise, where λ is the loss ratio ℓ_1/ℓ_2 with ℓ_1 indicating the loss associated with $x < c^*$ and β and ℓ_2 with $x \geq c^*$ and α (see, e.g., Emrick & Adams, 1969).

Now the problem is to estimate β and μ as well as the critical value c^* based upon these from a sample $x_1, \ldots, x_m$.

In our example, the mastery or random guessing model, (1) is taken to represent the probability of a number right score $X = x$ on a test of n items for an examinee randomly drawn from a population with a proportion of masters equal to μ. For a master the probability of a successful reply to an item is equal to β, whereas it is assumed that nonmasters guess blindly and have a probability of success equal to $\alpha = q^{-1}$, q being the number of item alternatives. The critical value c^* is used as a cut-off score on the test for classifying examinees as a master or a nonmaster. Since the quantities β, μ, and c^* are unknown, they must be estimated from the test scores of m randomly selected examinees.

Generally, the method of moments expresses the parameters to be estimated as explicit functions of population moments and next substitutes the corresponding sample moments to obtain estimators for these parameters. The method of moments ordinarily yields simple, closed-form estimators, which are consistent under mild conditions (Rao, 1972, p. 351).

Following Muench (1938) and Blischke (1962), we will not consider central moments or moment about the origin but the functions

$$\psi_k = E\ X(X - 1)\ldots(X - k + 1)\quad n(n - 1)\ldots(n - k + 1)$$

which are, up to the factor $n(n - 1)\ldots(n - k : 1)$, equal to the kth factorial moment and have the elegant property that for the mixture of binomials given in (1)

$$\psi_k = (1 - \mu)\alpha^k + \mu\beta^k. \tag{5}$$

Since there are only two parameters to be estimated, we can restrict our attention to $k = 1, 2$. For these values, it holds that

$$\psi_1 - \alpha = \mu(\beta - \alpha), \tag{6}$$

and

$$\psi_2 - \alpha^2 = \mu(\beta^2 - \alpha^2), \tag{7}$$

respectively. From (6), it follows that

$$\mu = \frac{\psi_1 - \alpha}{\beta - \alpha}, \tag{8}$$

while division of (6) into (7) yields

$$(9) \qquad \beta = \frac{\psi_2 - \alpha^2}{\psi_1 - \alpha} - \alpha$$

Moment estimates for β are now obtained by estimating ψ_1 and ψ_2 from a sample $x_1, \ldots, x_i, \ldots, x_m$ as

$$\hat{\psi}_1 = \frac{1}{m} \sum_{i=1}^{n} x_i/n = \bar{x}/n$$

$$\hat{\psi}_2 = \frac{1}{m} \sum_{i=1}^{n} x_i(x_i - 1)/n(n - 1) = s^2 + \bar{x}(\bar{x} - 1) \quad n(n - 1),$$

$\bar{x}$ and s^2 being the sample mean and variance, respectively, and substituting these into (9). Estimates for μ follow upon substitution of $\hat{\psi}_1$ and $\hat{\beta}$ into (8), Finally, $\hat{\beta}$ and $\hat{\mu}$ can be substituted into (4) to obtain an estimate of c^*.

Monte Carlo Results

The properties of the estimators derived in the preceding section were examined by means of a Monte Carlo experiment in which the two success parameters, α and β, the mixing parameter, μ, the test length, n, and the sample size, m, were varied and the consequences for the expected error of estimation and the risk function were determined. The last two were computed for both $\hat{\beta}$ and $\hat{\mu}$ as well as the estimated Bayes rule $\hat{c}^*$. The present experiment was basically a replication of an experiment used to investigate the properties of moment estimators for the case both success parameters as well as the mixing parameter are unknown (van der Linden, 1980b). The same parameter sets were used to be able to compare the outcomes of both experiments.

In Tables 1 - 4, ε is a generic symbol for an error of estimation and $E\varepsilon$ and $E\varepsilon^2$ denote the estimated expected error of estimation and risk function under squared error loss, respectively. The results reported in these tables are each based on 1,000 replications and data generated with the aid of random procedures from the NAG Fortran Library (1977). The center colum in Tables 1 - 3 represents the same parameter set and can be used as a benchmark.

Table 1 shows how the expected error of estimation and the risk vary

Insert Table 1 about here

as a function of (α, β). On the whole the results for $\hat{\beta}$ and $\hat{\mu}$ are excellent. All values of $E\varepsilon$ and $E\varepsilon^2$ differ only in their third decimal from zero, with the exception of those for $\hat{\mu}$ and (α, β) = (.40, .60), which are slightly larger. The values for $\hat{c}^*$, which were, as indicated by their subscripts, computed for λ equal to .25, 1, and 4, display the same pattern. For (.10, .90) and (.25, .75) the results are excellent, but a less favorable impression was obtained for (.40, .60). In this case the values for the risk function are too large, even when evaluated against n = 10. When (α, β) = (.40, .60), the situation approaches that of a single binomial. This expresses itself in less efficiency for $\hat{c}^*$ and, to some extent, for $\hat{\mu}$, indicating that the mixing parameter and the rule for seperating observations from both separate binomial distributions are estimated less reliably. The estimator of β is not influenced by the size of the difference between both success parameters, however.

In Table 2 it can be seen how the properties of the moment estimators

Insert Table 2 about here

depend upon the value of the mixing parameter μ. For all values of μ, the results for $\hat{\beta}$ and $\hat{\mu}$ as well as $\hat{c}^*$ are extremely good. (Again, it should be noted that the results for $\hat{c}^*$ must be evaluated on a scale differing by a factor of $n = 10$ from the one for $\hat{\beta}$ and $\hat{\mu}$.) The only tendency to be observed is a slight increase in the values of $E\varepsilon$ and $E\varepsilon^2$ with the value of μ. This is not surprising, since the larger the deviation of μ from .50, the more a situation of a single binomial is approached. Unlike Table 1, the moment estimator of μ itself does not display this tendency in Table 2.

Table 3 gives the results of our Monte Carlo experiment for test

Insert Table 3 about here

lengths of $n = 5$, $n = 10$, and $n = 20$. The results for $\hat{\beta}$ and $\hat{\mu}$ are equally good for all values of n used. Taking the scale of $\hat{c}^*$ into account, it is clear that the results for $\hat{c}^*$, already extremely good for $n = 5$, are better the larger the value of n.

From Table 4 it can be derived that even for a sample size of $m = 25$ $\hat{\beta}$

Insert Table 4 about here

and $\hat{\mu}$ have favorable properties. As for the expected error of estimation the same applies to $\hat{c}^*$, but the efficiency of $\hat{c}^*$ is less favorable compared with most of the results in Tables 1 - 3. As could be expected, an increase of the sample size pays well. Table 4 clearly shows how the values of $E\varepsilon$ and $E\varepsilon^2$ decrease with an increase in the value of m.

During the experiment we kept track of the number of cases inadmissible values for the moment estimators were met. No such cases arose, however. Since the values for α, β, μ, n, and m chosen in our experiment display a large variability, this seems to suggest that when using the moment estimators from this paper we do not need to worry about the possibility of meeting estimates lying outside the range of values their parameters are defined on.

Discussion

Comparing the above results with those obtained earlier for moment estimators for the case α is unknown (van der Linden, 1980b), it appears that knowing the true value of α yields considerable advantages. This holds for both $E\varepsilon$ and $E\varepsilon^2$, but most for the latter. In most cases a gain in efficiency equal to a factor of two or three can be observed. It also appears that the present estimators get less "upset" when the model comes close to a single binomial. Especially, the $E\varepsilon$ and $E\varepsilon^2$ values for $\hat{\beta}$ and $\hat{\mu}$ are hardly, if at all, influenced when this takes place; the values for $\hat{c}^*$ are in some cases low enough for use in practice. A comparable impression can be derived from a comparison of both sets of moment estimators for small sample sizes.

The cause of these differences between both sets of estimators goes back to the fact that the present estimators use only the first two sample moments, whereas the estimators for the case α is unknown are based on the third sample moment as well. This is more dependent on the relatively few observations in the tails of the sample distribution and it is therefore a less reliable estimate of the corresponding population moment.

The results in this paper can also be compared with results that were obtained for so-called endpoint estimators (van der Linden, 1980b). This method, which, to our knowledge, was introduced by Muench (1936) and used for Emrick's and Adam's model by Reulecke (1977a, 1977b), estimates the parameters from the frequencies in the tails of the sample distributions. When α is known, this method simply ignores the estimation of α and the estimators for β and μ remain the same. So the properties of $\hat{\beta}$ and $\hat{\mu}$ in this paper can be compared directly with those of the endpoint estimators for β and μ in van der Linden (1980b). However, the latter turned out to be beaten in all respects by the moment estimators for α unknown, and since these are in turn beaten by the estimators in the present paper, it follows that the last ones are always to be preferred when α is known.

It should be noted that the choice between the moment estimators for α known and α unknown does not depend only on whether or not a model is available which allows us to specify the value of α. The accuracy with which this can be done should also be taken into account. The estimators for α unknown have properties which are, although surpassed in quality by those for α known, excellent enough for most applications. When the error in specifying the value of α can be expected to be large, it might therefore be prudent to consider α an unknown parameter and to estimate accordingly. A comparison between the results presented in this paper and in van der Linden (1980b) can assist in this choice.

As a final comment we observe that the applicability of the model in this paper depends not only on the accuracy with which α can be specified but also on the degree to which the data can be approximated by a mixture of two binomials. In order to determine the latter, a straighforward procedure is to fit (1) to the sample distribution and to use the residuals for an ordinary χ^2-analysis. If the model does not appear to fit satisfactorily, the residuals can be inspected for possible improvements on the model. It is recommended that this be done when the model cannot be assumed to hold a priori.

References

Besel, R. Using group performance to interpret individual responses to criterion-referenced tests. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, Louisiana, February - March 1973. (EDRS No. ED 076 658).

Blischke, W.R. Moment estimators for the parameters of a mixture of two binomial distributions. Annals of Mathematical Statistics, 1962, 33, 444-454.

Blischke, W.R. Mixtures of discrete distributions. In Proceedings of the International Symposium on Discrete Distributions, Montreal, 1963.

Dayton, C.M., & Macready, G.B. A probabilistic model for validation of behavioral hierarchies. Psychometrika, 1976, 41, 189-204.

Emrick, J.A. An evaluation model for mastery testing. Journal of Educational Measurement, 1971, 8, 321-326.

Emrick, J.A., & Adams, E.N. An evaluation model for individualized instruction (Report RC 2674). Yorktown Hts., N.Y.: IBM Thomas J. Watson Research Center, October 1969.

Keats, J.A., & Lord, F.M. A theoretical distribution for mental test scores. Psychometrika, 1962, 27, 59-72.

Macready, G.B., & Dayton, C.M. The use of probabilistic models in the assessment of mastery. Journal of Educational Statistics, 1977, 2, 99-120.

Meskauskas, J.A. Evaluation models for criterion-referenced testing: Views regarding mastery and standard-setting. Review of Educational Research, 1976, 46, 133-158.

Muench, H. Probability distributions of protection test results. Journal of the American Statistical Association, 1936, 31, 677-689.

Muench, H. Discrete frequency distributions arising from mixtures of several single probability values. Journal of the American Statistical Association, 1938, 33, 390-398.

NAG Fortran Library Manual - Mark 6. Oxford, England: Numerical Algorithms Group, 1977.

Powers, L. Gene analysis by the partitioning method when interaction of genes are involved. Botanical Gazette, 1951, 113, 1-23.

Rao, C.R. The utilization of multiple measurements in problems of biological classification. Journal of the Royal Statistical Society, 1948, 10B, 159-203.

Rao, C.R. Linear statistical inference and its applications. New York: Wiley, 1973.

Reulecke, W.A. Ein Modell für die kriteriumorientierte Testauswertung. Zeitschrift für Empirische Pädagogik, 1977, 1, 49-72. (a).

Reulecke, W.A. A statistical analysis of deterministic theories. In H. Spada & W.F. Kempf (Eds.), Structural models of thinking and learning. Bern, Switzerland: Hans Huber Publishers, 1977. (b)

van der Linden, W.J. Decision models for use with criterion-referenced measurements. Applied Psychological Measurement, 1980, 4, in press. (a)

van der Linden, W.J. Estimating the parameters of Emrick's mastery testing model. Submitted for publication, 1980. (b)

Walsh, J.E. Approximate probability values for observed number of successes from statistically independent binomial events with unequal probabilities. Sankhyā, Series A, 1953, 15, 281-290.

Walsh, J.E. Definition and use of generalized percentage points. Sankhyā, Series A, 1959, 21, 281-288.

Walsh, J.E. Corrections to two papers concerned with binomial events. Sankhyā, Series A, 1963, 25, 247.

TABLE 1

Results for Moment Estimators with Varying Success Parameters and μ = .70, n = 10, and m = 100

(α, β)	(.10, .90)		(.25, .75)		(.40, .60)	
	$E\varepsilon$	$E\varepsilon^2$	$E\varepsilon$	$E\varepsilon^2$	$E\varepsilon$	$E\varepsilon^2$
$\hat{\beta}$	.000	.000	-.001	.000	-.002	.001
$\hat{\mu}$	.002	.002	.001	.003	.024	.020
$\hat{c}^*_{.25}$	-.003	.024	-.013	.039	-.044	2.010
$\hat{c}^*_1$	-.003	.027	-.015	.051	-.510	2.925
$\hat{c}^*_4$	-.003	.031	-.017	.064	-.584	4.102

TABLE 2

Results for Moment Estimators with Varying Mixing

Parameter and α = .25, β = .75, n = 10, and m = 100

μ	.50		.70		.90	
	$E\varepsilon$	$E\varepsilon^2$	$E\varepsilon$	$E\varepsilon^2$	$E\varepsilon$	$E\varepsilon^2$
$\hat{\beta}$	.000	.001	-.001	.000	.000	.000
$\hat{\mu}$	.000	.004	.001	.003	.000	.001
$\hat{c}^*_{.25}$	.002	.048	-.013	.039	-.004	.079
$\hat{c}^*_1$	.001	.067	-.015	.051	-.037	.088
$\hat{c}^*_4$	.000	.090	-.017	.064	-.038	.099

TABLE 3

Results for Moment Estimators with Varying Test

Length and α = .25, β = .75, μ = .70, and m = 100

n	5		10		20	
	$E\varepsilon$	$E\varepsilon^2$	$E\varepsilon$	$E\varepsilon^2$	$E\varepsilon$	$E\varepsilon^2$
$\hat{\beta}$	.001	.001	-.001	.000	.000	.000
$\hat{\mu}$	.000	.004	.001	.003	-.001	.003
$\hat{c}^*_{.25}$	-.008	.039	-.013	.039	.004	.050
$\hat{c}^*_1$	-.009	.056	-.015	.051	.005	.059
$\hat{c}^*_4$	-.010	.078	-.017	.064	.005	.069

TABLE 4

Results for Moment Estimators with Varying Sample

Size and α = .25, β = .75, μ = .70, and n = 10

m	25		50		500	
	$E\varepsilon$	$E\varepsilon^2$	$E\varepsilon$	$E\varepsilon^2$	$E\varepsilon$	$E\varepsilon^2$
$\hat{\beta}$	.002	.002	-.001	.001	.000	.000
$\hat{\mu}$	.005	.012	.002	.007	.000	.001
$\hat{c}^*_{.25}$	-.031	.192	-.023	.087	.000	.007
$\hat{c}^*_1$	-.031	.245	-.026	.112	-.001	.010
$\hat{c}^*_1$	-.032	.308	-.029	.142	-.001	.012

Acknowledgement

The author is indebted to Gideon J. Mellenbergh and Ivo W. Molenaar for their comments on the first draft of this paper.

Authpr's address

Wim J. van der Linden, Onderafdeling Toegepaste Onderwijskunde, Technische Hogeschool Twente, Postbus 217, 7500 AE Enschede, The Netherlands.